

One-step approximations for detecting regime changes in the state space model with application to the influenza data

Tianni Zhou^{a,*}, Robert Shumway^b

^a*Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA*

^b*Department of Statistics, University of California, Davis, CA 95616, USA*

Received 16 July 2007; received in revised form 20 August 2007; accepted 20 August 2007

Available online 12 September 2007

Abstract

Kim and Nelson [1999. *State Space Models with Regime Switching*. MIT Press, Cambridge, MA] and others extended the framework of state space models involving independent regime changes to the Markov dependent case. The cost of dealing with state space models with Markov switching is high in computational effort because of the number of the possible paths through the chain. Thus it is necessary to make some approximations in order to obtain a computationally feasible algorithm for estimation. The approximations depend on modified smoothing and filtering recursions that can be easily incorporated into an EM algorithm for maximum likelihood estimation. To investigate the accuracy of approximations, we develop a new method to obtain more exact solutions, and then compare the two methods. We apply both methods to a simulated series. The result shows that employing the approximation-based algorithm not only provides accurate results but also leads to a significant reduction in the computational costs. We also apply the methods to an influenza mortality series, in which we develop a model that is general enough to include most structural models useful in monitoring changes of regime. The model proposed has the flexibility to deal with a wide range of problems involving possible regime shifts in pattern that may be seen to occur in many biological, medical and epidemiological studies.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Hidden Markov model; State space model with switching; Kalman filter; Kalman smoother; Markov chain Monte Carlo; Gibbs sampling; EM algorithm; Influenza

1. Introduction

The state space model has become a powerful tool for modeling and forecasting dynamic systems. A growing number of published papers that employ it demonstrate its usefulness and broadness of application. In essence, a state space model is one in which an observed variable is the sum of a linear function of the state variable and an error. The state variable, in turn, evolves according to a stochastic difference equation that depends on parameters that are generally unknown. Hence, for such a state space model, given the observed data, estimations of the unobserved state vector and the parameters are of primary interest.

A switching state space model is obtained, if we assume that, in addition to the unobserved state variable, a discrete unknown switching variable influences the distribution of the observed data. A typical way of including a switching

* Corresponding author.

E-mail address: tzhou@usc.edu (T. Zhou).

mechanism into a Gaussian state space model is to assume that one of the variances, e.g. the variance appearing in the

Note that s_t does not affect the prior state x_{t-1} and observed data Y_{t-1} . Conditional on $s_t = j$, the Kalman filter produces predictions $x_{t|t-1}^{(j)} = \phi x_{t-1|t-1}$ and mean squared error matrices $P_{t|t-1}^{(j)} = \phi P_{t-1|t-1} \phi' + Q$ for all $j = 1, \dots, M$. Hence,

$$x_{t|t-1} = \phi x_{t-1|t-1} \quad (2)$$

and

$$P_{t|t-1} = \phi P_{t-1|t-1} \phi' + Q. \quad (3)$$

For model (1), given $s_t = j$, the Kalman filter gives the updated inferences $x_{t|t}^{(j)}$ and $P_{t|t}^{(j)}$:

$$x_{t|t}^{(j)} = x_{t|t-1} + K_{t|t-1}^{(j)}(y_t - A_j x_{t|t-1}), \quad (4)$$

and

$$P_{t|t}^{(j)} = P_{t|t-1} - K_{t|t-1}^{(j)} A_j P_{t|t-1}, \quad (5)$$

where

$$K_{t|t-1}^{(j)} = P_{t|t-1} A_j' (A_j P_{t|t-1} A_j' + R)^{-1}$$

is the Kalman gain. Incorporating the filtered probabilities $\pi_j(t|t) = P[s_t = j|Y_t]$ into $x_{t|t}^{(j)}$ and $P_{t|t}^{(j)}$, we obtain

$$x_{t|t} = x_{t|t-1} + \sum_{j=1}^M \pi_j(t|t) K_{t|t-1}^{(j)} (y_t - A_j x_{t|t-1}) \quad (6)$$

and

$$P_{t|t} = \sum_{j=1}^M \pi_j(t|t) (P_{t|t}^{(j)} + (x_{t|t}^{(j)} - x_{t|t})(x_{t|t}^{(j)} - x_{t|t})'). \quad (7)$$

Note that a modified term $(x_{t|t}^{(j)} - x_{t|t})(x_{t|t}^{(j)} - x_{t|t})'$ is introduced in (7) due to condition on $s_t = j$.

To implement the filtering algorithm, we need to make inferences on the filtered probabilities $\pi_j(t|t) = P(s_t = j|Y_t)$. From the theory of hidden Markov chain it is well known that

$$\pi_j(t|t) = \frac{f_j(t|t-1)\pi_j(t|t-1)}{\sum_{k=1}^M f_k(t|t-1)\pi_k(t|t-1)}, \quad (8)$$

where $f_j(t|t-1) = f(y_t|s_t = j, Y_{t-1})$ and $\pi_j(t|t-1) = P(s_t = j|Y_{t-1}) = \sum_{i=1}^M \pi_i(t-1|t-1)\pi_{ij}$ (Shumway and Katzoff, 1992).

It remains to calculate the filter density $f(y_t|s_t = j, Y_{t-1})$. The difficulty in calculating the density is the dependence among the y_t 's, which means it is necessary to enumerate over all possible histories of s_t 's to calculate the density. It turns out that $f(y_t|s_t = j, Y_{t-1})$ is a mixture of Gaussian distributions.

Since the evaluation of $f(y_t|s_t = j, Y_{t-1})$ is highly computer intensive with the exponential growth complexity, an approximation is introduced in making inferences on the filtered probabilities, that is, we approximate $f(y_t|s_t = j, Y_{t-1})$ by a normal distribution with the mean $A_j x_{t|t-1}^{(j)}$ and variance $A_j P_{t|t-1}^{(j)} A_j' + R$, as in Smith and West (1983) and Shumway and Stoffer (1991).

Using the obtained filtered probabilities, we may calculate the smoothed probabilities $\pi_{jk}(t+1|n) = P(s_t = j, s_{t+1} = k|Y_n)$ and $\pi_j(t|n) = P(s_t = j|Y_n)$. First, note that for model (1), if s_{t+1} is known, then the information on s_t will not affect $y_{t+1}, \dots, y_n$. Hence, $P(s_t = j|s_{t+1} = k, Y_n) = P(s_t = j|s_{t+1} = k, Y_t)$. It follows that

$$\pi_{jk}(t+1|n) = P(s_t = j, s_{t+1} = k|Y_n) = \frac{\pi_k(t+1|n)\pi_{jk}\pi_j(t|t)}{\pi_k(t+1|t)} \quad (9)$$

and

$$\pi_j(t|n) = \sum_{k=1}^M \pi_{jk}(t+1|n). \quad (10)$$

The smoothed probabilities $\pi_j(t|n)$ and $\pi_{jk}(t+1|n)$ are calculated by backward recursion.

We now consider the problem of obtaining inferences on x_t and s_t using all the information in the sample. The Kalman smoother yields

$$x_{t|n}^{(j)} = x_{t|t}^{(j)} + J_t^{(j)}(x_{t+1|n} - x_{t+1|t}), \quad (11)$$

and

$$P_{t|n}^{(j)} = P_{t|t}^{(j)} + J_t^{(j)}(P_{t+1|n} - P_{t+1|t})J_t^{(j)'} \quad (12)$$

where $J_t^{(j)} = P_{t|t}^{(j)}\phi'(P_{t+1|t})^{-1}$. It is worth noting that because x_t, Y_n lack joint normality, these two formulas only give approximations to $x_{t|n}^{(j)}$ and $P_{t|n}^{(j)}$ for t

4. For $t = n, n - 1, \dots, 1$, use (9) and (10) to calculate the smoothed probabilities. Use the Kalman smoother (11) and (12) to calculate the smoothed expectations and covariances of x_t , given $s_t = j$, $j = 1, 2, \dots, M$, use (13) and (14) to incorporate the smoothed probabilities into $x_{t|n}^{(j)}$ and $P_{t|n}^{(j)}$ and calculate the smoothed expectations $x_{t|n}$ and covariances $P_{t|n}$, and use (15) to calculate the lag-one covariances.
5. Update the estimates μ_0 , ϕ_0 , Q , R , and π_j , $\pi_{i,j}$, $i, j = 1, 2, \dots, M$, using (A.2)–(A.7).
6. Repeat steps 2–5 to convergence.

3. Monte Carlo expectation maximization (MCEM)

To investigate the accuracy of approximations used in the AEM algorithm, we need to compare the approximate smoothing in AEM with a Monte Carlo-based version. Monte Carlo integration techniques such as Gibbs sampler (Carlin et al., 1992; Carter and Kohn, 1994; Kim and Nelson, 1999) are useful in such a comparison. Based on their approaches, we develop an MCEM to obtain more exact solutions. A justification of the EM algorithm based on approximations in the expectation or maximization steps is given by Neal and Hilton (1998).

MCEM contains an inner loop and an outer loop. The inner loop uses the Gibbs sampling approach to generate the state variables and the switching variables. Note that this step does not use any approximation, and we can obtain exact filters and smoothers. Based on the exact filters and smoothers, we then run EM to estimate the parameters. The details of MCEM algorithm are given below.

Step 1: Generate $X_n = [x_1 \ x_2 \ \dots \ x_n]$, conditional on S_n , Y_n , and hyperparameters of the model.

Step 2: Generate $S_n = [s_1, \dots, s_n]$, conditional on the observed data, the state variables, and hyperparameters of the model using Gibbs Sampling.

Step 3: Update the initial and transition probabilities, given the observed data and the state variables.

Step 4: Based on the exact filters and smoothers, estimate the parameters using EM.

In the first step, the state variables $X_n = [x_1 \ x_2 \ \dots \ x_t]$ are generated from the joint distribution given by

$$p(X_n|Y_n) = p(x_n|Y_n) \prod_{t=1}^{n-1} p(x_t|x_{t+1}, Y_t). \quad (16)$$

Eq. (16) suggests that the whole vector of X_n can be generated by first generating x_n from $p(x_n|Y_n)$, and then, for $t = n - 1, n - 2, \dots, 2, 1$, generating x_t from $p(x_t|x_{t+1}, Y_t)$, given the generated values for x_{t+1} . Note that the state space model conditional on $S_n = [s_1, s_2, \dots, s_n]$ is linear and Gaussian. The distribution of x_n given Y_n and that of x_t given x_{t+1} and Y_t for $t = n - 1, \dots, 1$ are also Gaussian. Hence, we may use the Kalman filter, without any approximation, to obtain

$$x_{n|n} = E(x_n|Y_n),$$

$$P_{n|n} = \text{cov}(x_n|Y_n),$$

$$x_{t|t, x_{t+1}} = E(x_t|Y_t, x_{t+1}) = E(x_t|x_{t+1}, Y_t),$$

$$P_{t|t, x_{t+1}} = \text{cov}(x_t|Y_t, x_{t+1}) = \text{cov}(x_t|x_{t+1}, Y_t).$$

Then we have

$$x_n|Y_n \sim N(x_{n|n}, P_{n|n})$$

and

$$x_t|Y_t, x_{t+1} \sim N(x_{t|t, x_{t+1}}, P_{t|t, x_{t+1}}).$$

By running the last iteration of the usual updating procedure of the Kalman filter, we obtain $x_{n|n}$ and $P_{n|n}$, given S_n , and hyperparameters of the model.

$x_{t|t, x_{t+1}}$ and $P_{t|t, x_{t+1}}$ are derived by updating an estimate of x_t based on a combination of $x_{t|t}$ and information contained in the generated x_{t+1} (Kim and Nelson, 1999) and then x_t , $t = n - 1, n - 2, \dots, 1$ are generated from $N(x_{t|t, x_{t+1}}, P_{t|t, x_{t+1}})$, where

$$x_{t|t, x_{t+1}} = x_{t|t} + P_{t|t} \phi' (\phi P_{t|t} \phi' + Q)^{-1} (x_{t+1} - \phi x_{t|t})$$

and

$$P_{t|t, x_{t+1}} = (P_{t|t}^{-1} + \phi' Q^{-1} \phi)^{-1}.$$

In the second step, we generate $S_n = [s_1, \dots, s_n]$ conditional on the observed data, the state variables, and hyperparameters of the model. Carter and Kohn (1994) showed how to generate the indicator variables, which are the variables used to show which members of the mixture each of the observation noise and signal noise belong to. Based on the algorithm in generating indicator variables, we generate S_t as follows. We omit the dependence on hyperparameters of the model and denote $X_t = [x_1, x_2, \dots, x_t]$ and $Y_t = [y_1, y_2, \dots, y_t]$.

In forward recursion, we calculate, for $t = 1, \dots, n$,

$$p(s_t = j | Y_t, X_t) = \frac{p(y_t | x_t, s_t = j) p(x_t | x_{t-1}, s_t = j) p(s_t = j | Y_{t-1}, X_{t-1})}{\sum_{j=1}^M p(y_t | x_t, s_t = j) p(x_t | x_{t-1}, s_t = j) p(s_t = j | Y_{t-1}, X_{t-1})} \quad (17)$$

and

$$p(s_{t+1} = k | Y_t, X_t) = \sum_{j=1}^M p(s_{t+1} = k | s_t = j) p(s_t = j | Y_t, X_t). \quad (18)$$

To generate S_n , we first generate s_n from $p(s_n | Y_n, X_n)$ and then for $t = n - 1, \dots, 1$, we generate s_t from $p(s_t = j | Y_t, X_t, s_{t+1} = k)$, where

$$p(s_t = j | Y_t, X_t, s_{t+1} = k) = \frac{p(s_{t+1} = k | s_t = j) p(s_t = j | Y_t, X_t)}{p(s_{t+1} = k | Y_t, X_t)}. \quad (19)$$

In the inner loop, we need to update the initial probabilities and transition probabilities. Using the filtered probabilities (17) and (18), we calculate the smoothed probabilities

$$P(s_t = j, s_{t+1} = k | Y_n, X_n) = \frac{P(s_{t+1} = k | Y_n, X_n) P(s_{t+1} = k | s_t = j) P(s_t = j | Y_t, X_n)}{P(s_{t+1} = k | Y_t, X_n)}$$

and

$$P(s_t = j | Y_n, X_n) = \sum_{k=1}^M P(s_t = j, s_{t+1} = k | Y_n, X_n).$$

The maximum likelihood estimators for the initial probabilities and transition probabilities are given by

$$\hat{P}(s_1 = j) = P[s_1 = j | Y_n, X_n],$$

$$\hat{P}(s_t = j | s_{t-1} = i) = \frac{\sum_{t=2}^n P(s_{t-1} = i, s_t = j | Y_n, X_n)}{\sum_{t=2}^n P(s_{t-1} = i | Y_n, X_n)}.$$

Note that these estimators are obtained by maximizing the expectation of the complete data likelihood function $f(x_1, \dots, x_n, y_1, \dots, y_n, s_1, \dots, s_n)$ conditional on Y_n and X_n .

The outer loop is completed by updating hyperparameters of the model, using the maximum likelihood estimators as in (A.4)–(A.7).

4. Applications to simulated data and modeling influenza data

4.1. Simulated data

We study the proposed algorithms by analyzing a simulated series. Assume that observations are generated by

$$y_t = A_{s_t} x_t + v_t, \quad (20)$$

where the state vector $x_t = (x_{t1}, x_{t-1,1}, x_{t2}, x_{t3})$ has components satisfying

$$\begin{aligned} x_{t1} &= 1.5x_{t-1,1} - 0.6x_{t-2,1} + w_{t1}, \\ x_{t2} &= 0.15x_{t-1,2} + x_{t-1,3} + w_{t2}, \\ x_{t3} &= x_{t-1,3}, \quad t = 1, \dots, 150. \end{aligned} \quad (21)$$

The underlying signal, x_{t1} , is always present and is modeled as the second-order autoregressive series. The process x_{t2} models a first-order AR(1) component plus a constant, which is sometimes present and sometimes absent. We assume that the measurement matrix, A_{s_t} , is dependent on s_t ($s_t = 1, 2$), a discrete random variable defined by the initial probabilities $\pi_j = P(s_1 = j)$, $j = 1, 2$ with $\sum_{j=1}^2 \pi_j = 1$, and the stationary transition probabilities $\pi_{ij} = P(s_t = j | s_{t-1} = i)$ with $\sum_{j=1}^2 \pi_{ij} = 1$ for all i . Here, A_{s_t} takes one of two possible forms $A_1 = [1, 0, 0]$ and $A_2 = [1, 1, 0]$. When $s_t = 1$, y_t is modeled as $y_t = x_{t1} + v_t$. When $s_t = 2$, y_t is modeled as $y_t = x_{t1} + x_{t2} + v_t$.

The model specified above can be written in the general state space form. The state equation is

$$\begin{pmatrix} x_{t1} \\ x_{t-1,1} \\ x_{t2} \\ x_{t3} \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0.15 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-2,1} \\ x_{t-1,2} \\ x_{t-1,3} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ 0 \\ w_{t2} \\ 0 \end{pmatrix}.$$

The observation equation is

$$y_t = A_{s_t} x_t + v_t, \quad t = 1, \dots, n.$$

The v_t are distributed as iid $N(0, R)$ random variables, and the w_{t1} are distributed as $N(0, q_{11})$ and w_{t2} are distributed as $N(0, q_{22})$ random variables, independent of the v_t , $t = 1, \dots, 150$. Note that x_{t3} stays constant and depends completely on the mean and the variance that determine the normal distribution $x_{03} \sim N(\mu_{03}, \sigma_{03}^2)$ of the initial state.

For this simulation, we use a sample size of 150 with $R = 0.01$, $q_{11} = 1$, $q_{22} = 0.1$, $\mu_{03} = 10$, $\sigma_{03}^2 = 0.01$, initial probabilities $\pi_1 = \pi_2 = 0.5$, and transition probabilities $\pi_{11} = \pi_{22} = 0.9$ ($\pi_{12} = \pi_{21} = 0.1$). For the AEM, we carry out the procedures in Sections 2.1 and 2.2 to obtain smoothers and parameter estimates. For MCEM, we use Gibbs sampling algorithm presented in Section 3 to generate the state variables and switching variables, and estimate parameters using the EM procedure outlined in Section 2.2.

Fig. 1 compares the smoother for the AR(2) component from the two methods. The smoother estimated from the AEM overlays that estimated from MCEM. Fig. 2 shows the estimated switching variable s_t obtained using the AEM. It is clear that the smoother indicates A_2 for $25 \leq t \leq 65$, $105 \leq t \leq 108$, $196 \leq t \leq 230$, $350 \leq t \leq 432$, and $440 \leq t \leq 475$, and A_1 for other time points. These are states exhibiting high probabilities $P(S_t = 2)$ over a significant portion of the range. The plot of the estimated switching variables s_t obtained from MCEM using Eqs. (17)–(19) overlays the estimated switching variables in Fig. 2. Inferences about the periods of switching are the same for the two methods.

Table 1 compares the estimated parameters ϕ_{11} , ϕ_{12} , R , q_{11} , q_{22} , π_{11} , and π_{22} obtained by the two methods. Results show that although MCEM provides more exact solutions, approximations of the AEM are also quite accurate. All parameters converged rather quickly. The iterations were considered to have converged when the maximum rate of change of the log-likelihood estimates between iteration steps was less than 0.005. For most of them, it took as few as 50 iterations to converge.

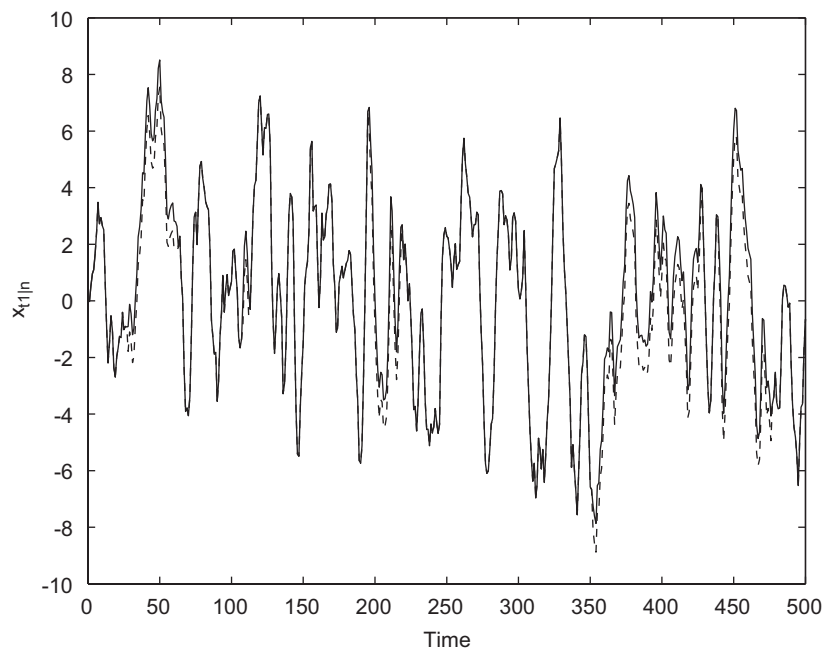


Fig. 1. Smoothers (AR2 component) from AEM (solid) and MCEM (dotted).

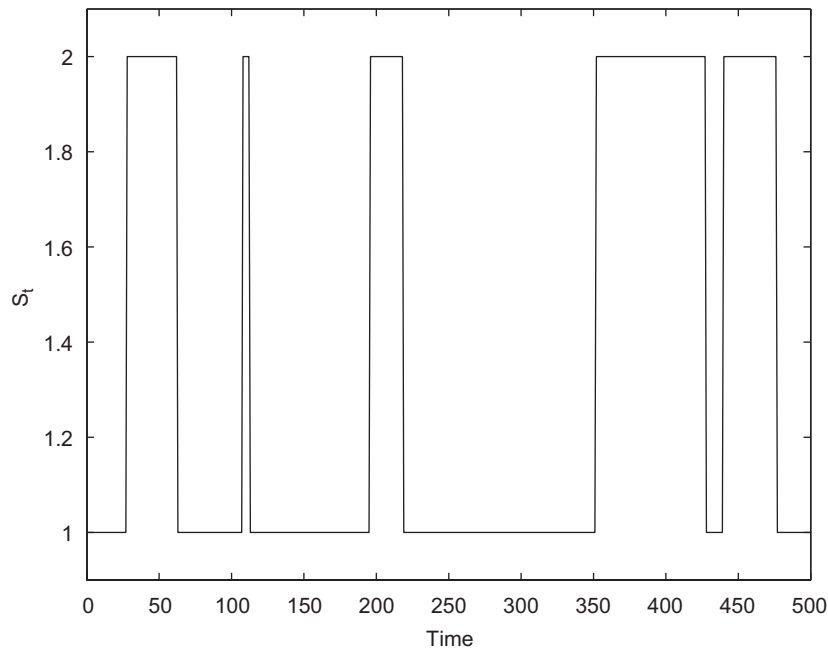


Fig. 2. Estimated switching variables s_t obtained using AEM/MCEM.

Time requirements for the MCEM were compared to the AEM. Since MCEM has an inner loop and an outer loop, it took much more time than did the AEM algorithm. For example, MCEM needed about 18 min to complete a run of 100(inner)/100(outer) iterations. On the other hand, the AEM algorithm took only 15 s to finish a run of 100 iterations.

Table 1
Comparisons of estimates from AEM and MCEM

	ϕ_{11}	ϕ_{12}	R	q_{11}	q_{22}	π_{11}	π_{22}
AEM	1.43	−0.57	0.0087	0.92	0.12	0.9167	0.8523
MCEM	1.46	−0.58	0.0091	1.01	0.09	0.9012	0.8774
True	1.5	−0.6	0.01	1	0.1	0.9	0.9

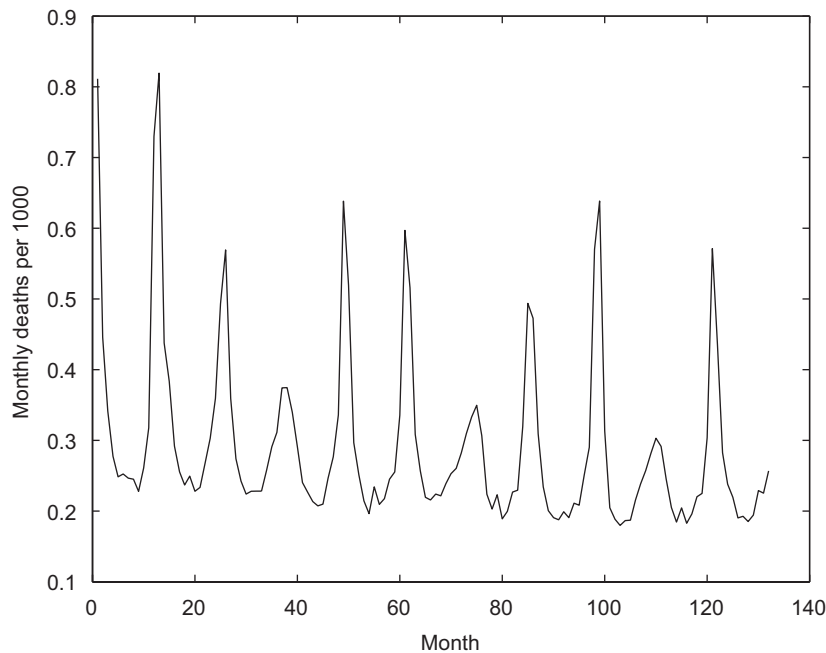


Fig. 3. Influenza data.

It is clear that tremendous reductions in computing time are available through AEM. The above computational-related comments were made using a computer with Pentium-IV 3 GHz CPU and Windows XP operating system.

4.2. Influenza data

To illustrate the methods of this paper, we apply the procedure to the influenza data. These data were analyzed by Shumway and Stoffer (2000). We apply our methodology to this data and compare our results with theirs at the end of this section. Fig. 3 shows a plot of monthly pneumonia and influenza deaths per 1000 in the US for 11 years, 1968–1978. Monthly pneumonia and influenza deaths do not follow a linear Gaussian process because there are large bursts of positive and negative changes that occur periodically in this series, probably resulting from influenza epidemics. Although the number of deaths is typically largest during the winter months, the data are not perfectly seasonal. That is, although the peak of the series often occurs in January, in other years, the peak occurs in December, February, or March. The change in the numbers of deaths slowly rises for some months and, then, sometime in the winter, has a possibility of jumping to a large number. If the processes do make a large jump, then a subsequent significant decrease occurs in flu deaths.

Let y_t denote the mortality caused by influenza at month t . We observe that there is a slight negative trend in the data. We first remove a fixed linear trend from the data and then model the data in terms of a structural component model coupled with a hidden Markov process that determines whether a flu epidemic exists.

Table 2

Comparisons of estimates from AEM and MCEM for influenza data

	ϕ_{11}	ϕ_{12}	R	σ_1^2	π_{11}	π_{12}
AEM	0.9088	−0.3971	0.001	0.0030	0.9570	0.6854
MCEM	0.8550	−0.3732	0.001	0.0038	0.9328	0.7162

The model consists of two structural components. The first component, x_{t1} , is an AR(2) process chosen to represent the periodic (seasonal) component of the data

$$x_{t1} = \phi_{11}x_{t-1,1} + \phi_{12}x_{t-2,1} + w_{t1},$$

where w_{t1} is white noise, with $\text{var}(w_{t1}) = \sigma_1^2$.

The second component, x_{t2} , is a random effect, which is chosen to represent the sharp rise in the data during an epidemic. This corresponds to a simple level shift during a flu epidemic.

$$x_{t2} = x_{t-1,2}.$$

Throughout the years, periods of normal influenza mortality (state 1) are modeled as $y_t = x_{t1} + v_t$, where the measurement error, v_t , is white noise with $\text{var}(v_t) = R$. When an epidemic occurs (state 2), mortality is modeled as $y_t = x_{t1} + x_{t2} + v_t$.

The model specified above can be written in the general state space form. The state equation is

$$\begin{pmatrix} x_{t1} \\ x_{t2} \\ x_{t-1,1} \end{pmatrix} = \begin{pmatrix} \phi_{11} & 0 & \phi_{12} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_{t-1,1} \\ x_{t-1,2} \\ x_{t-2,1} \end{pmatrix} + \begin{pmatrix} w_{t1} \\ 0 \\ 0 \end{pmatrix}.$$

The observation equation is

$$y_t = A_{s_t}x_t + v_t, \quad t = 1, \dots, n,$$

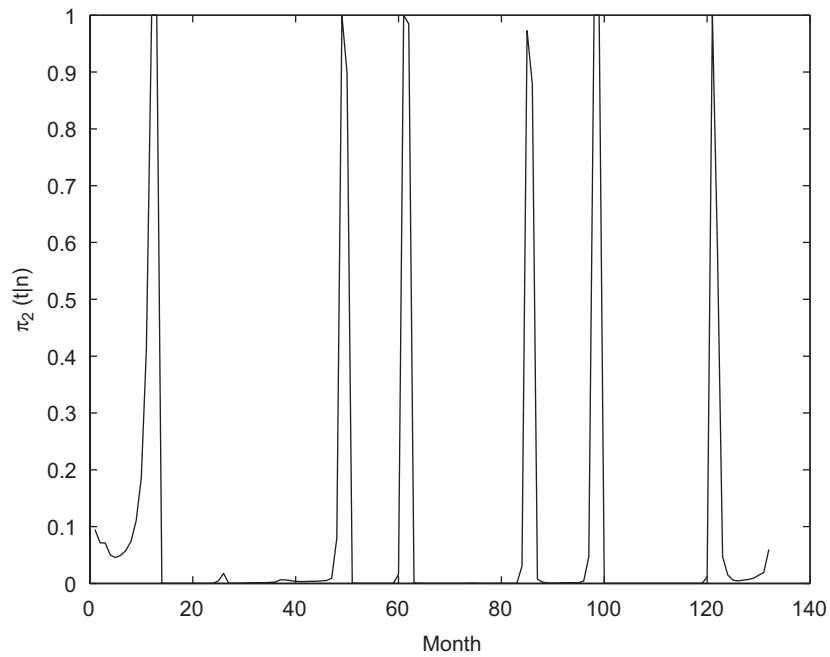
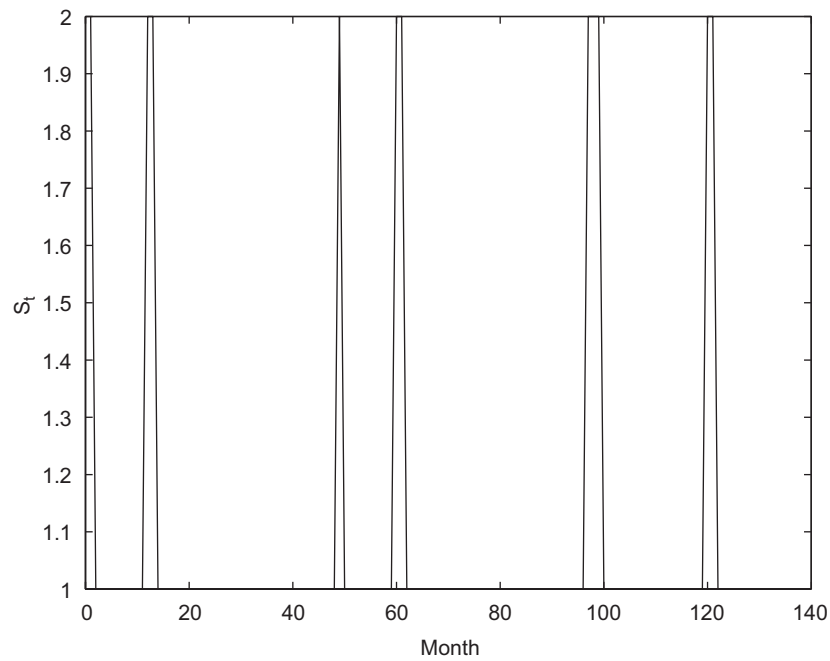
where $s_t \in \{1, 2\}$, $x_t = [x_{t1}, x_{t2}, x_{t-1,1}]'$, $A_1 = [1, 0, 0]$ (non-epidemic), and $A_2 = [1, 1, 0]$ (epidemic).

We assume that A_{s_t} is generated by a hidden Markov chain, defined by s_t . We use initial probabilities $\pi_1 = 0.5$ and $\pi_2 = 0.5$. From the influenza data, we observe that there are around six or seven substantial peaks out of 132 observation points and use transition probability $\frac{7}{132} = 0.05$.

The initial state vector was set to $\mu = (0, 0.4, 0)$. We ran the AEM and MCEM. It seemed that a few parameters converge rather slowly. We tried to solve the problem by setting some parameters to reasonable fixed values. We noticed that, after running some iterations of both procedures, the variance of the observation noise, R , became small, close to 0. We thus fixed $R = 0.001$ and estimated other parameters. The small variance R of the measurement errors tells us that the variability in data is explained solely by the variance components of the state system. We also tried to update the initial mean x_0 but the estimation became unstable for MCEM. Using EM, we estimated the initial mean as $x_0 = (0.012, 0.321, 0.002)$, and then fixed the initial mean at this value in MCEM. Table 2 compares the final estimates obtained by running the AEM algorithm and MCEM.

Figs. 4 and 5 show the estimated switching variable s_t , $t = 1, \dots, n$ obtained by AEM and MCEM. $s_t = 1$ corresponds to the period of normal influenza mortality and $s_t = 2$ corresponds to the period of flu epidemic. The estimated s_t values seem to be on the mark. With AEM, six epidemic periods, 1968, 1972, 1973, 1975, 1976, and 1978 were detected, each corresponding to a peak in the observed influenza data. The peak around the year of 1970 (month 25), seems noticeable by eye, however was not detected by the algorithm. The peaks at year of 1970 (month 25) and 1975 (month 85) were not detected by MCEM.

Following results are obtained using the AEM. Fig. 6 shows the smoothed structural components of influenza mortality. Note that there is a simple level shift during a flu epidemic. The first component represents the seasonal aspect of the data, and the constant represents the sharp rise during a flu epidemic. The estimated constant is 0.2229, suggesting an increase in mortality by about 0.2229 per 1000 during a flu epidemic.

Fig. 4. Estimated switching variables s_t (AEM).Fig. 5. Estimated switching variables s_t (MCEM).

The predicted y_t is obtained as follows:

$$y_t = \hat{x}_{t1} \quad \text{if } \pi_1(t|n) > \pi_2(t|n),$$

$$y_t = \hat{x}_{t1} + \hat{x}_{t2} \quad \text{if } \pi_1(t|n) \leq \pi_2(t|n),$$

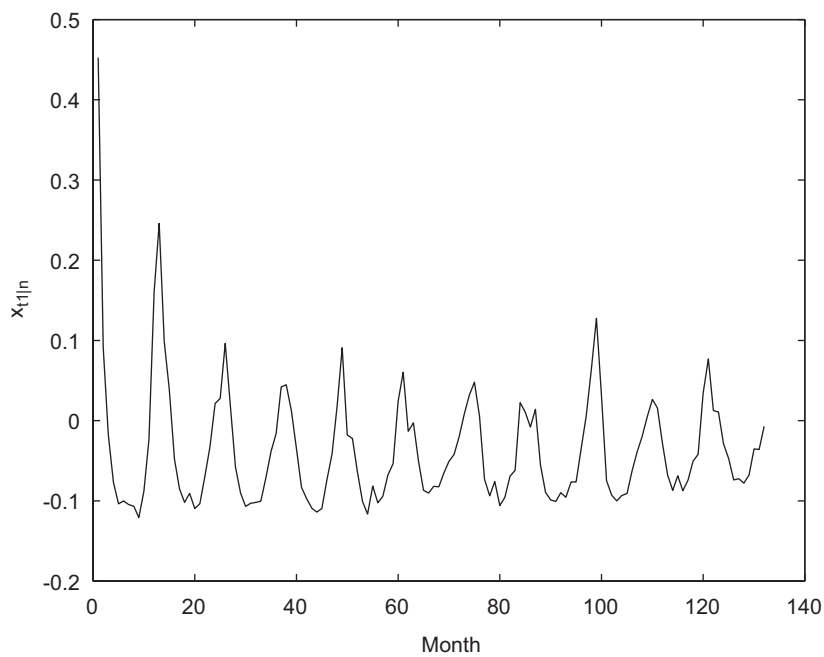


Fig. 6. Smoothed structural component: cyclic trace.

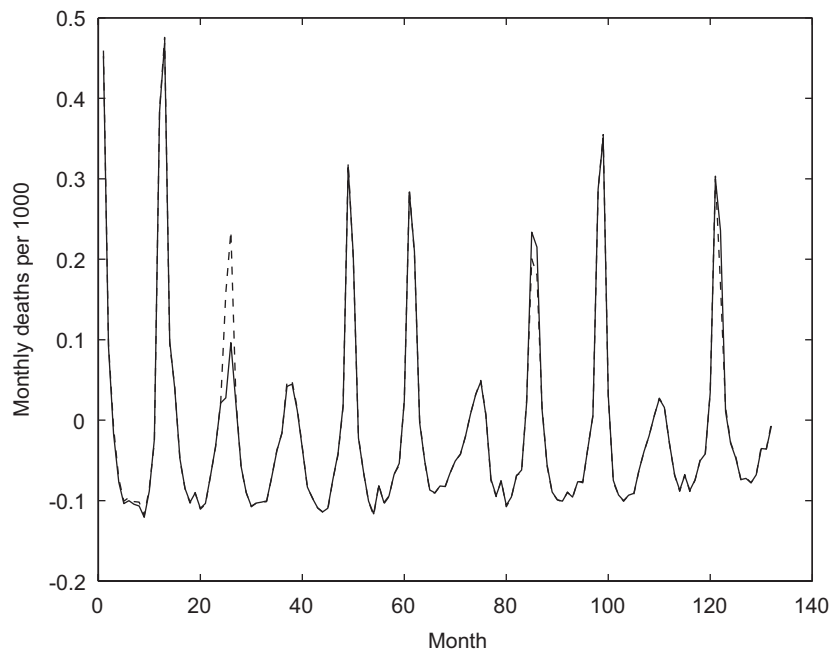


Fig. 7. Transformed y_t (dotted) vs. predicted y_t (solid).

where $\hat{x}_{t1}$ and $\hat{x}_{t2}$ are the smoothed estimators of the two structural components. Fig. 7 shows the observed mortality with linear trend removed and the predicted mortality. Fig. 8 shows the observed mortality and the predicted mortality with the added fixed linear trend. The predicted mortality is identical to the true values, except that there is an underestimation in 1970.

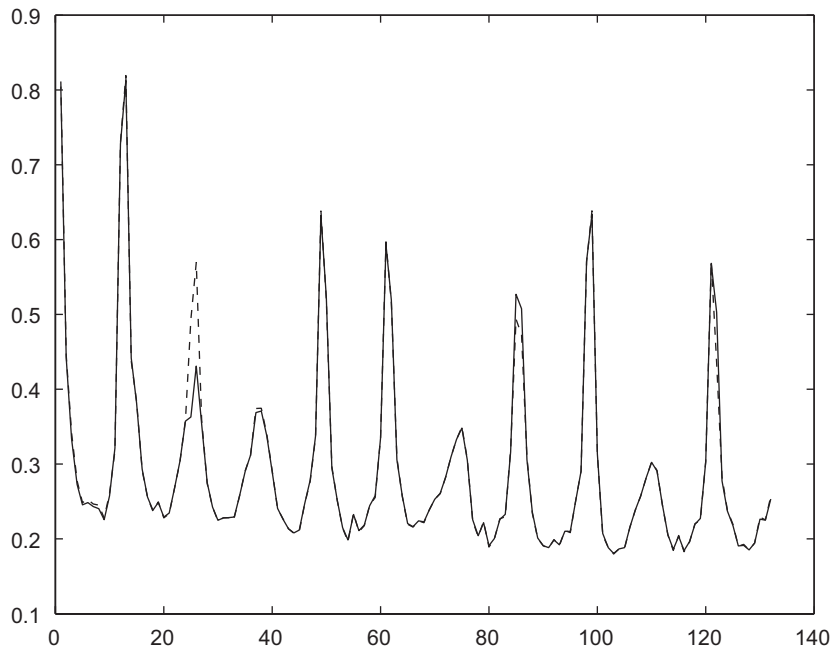


Fig. 8. Observed y_t (dotted) vs. predicted y_t adding linear trend (solid).

Shumway and Stoffer used a different model to fit this data. Their model consists of three structural components. The first component is an AR(2) process chosen to represent the seasonal component of the data. The second component is an AR(1) process with a nonzero constant term, which is chosen to represent the sharp rise in the data during an epidemic. The third component is a fixed trend component. They applied the approximate filtering algorithm to obtain the filtered structural components but did not obtain the smoothed structural components of the influenza mortality. Shumway and Stoffer estimated parameters using a quasi-Newton–Raphson procedure. Their results also showed that there is very little measurement error, and the variability in data is explained solely by the variance components of the state system. Their results and ours agreed on the fact that there is an increase in mortality by about 0.2 per 1000 during a flu epidemic. Our results showed a smaller variance components in the state system than theirs. Shumway and Stoffer obtained one-month-ahead prediction and their results showed that the predicted mortality tends to be underestimated during five epidemic periods. As seen in Fig. 8, we have underestimation at only one epidemic period. We obtained the predicted mortality using smoothed probabilities and smoothed structural components. Comparing to Shumway and Stoffer’s model, ours seems simpler and improves the prediction.

5. Conclusions

In this paper, we have developed and extended the framework of state space models considered by Shumway and Stoffer (1991), and demonstrated how these models may be used to handle regime shifts in longitudinal biological and medical time series. Shumway and Stoffer (1991) developed a methodology for modeling state space models involving independent regime changes and Kim and Nelson (1999) and others extended it to the Markov-dependent case. The EM algorithm was applied to the general state space model without switching by Shumway and Stoffer (1982) and to the state space model with independent switching by Shumway and Stoffer (1991). Our updating equations for the parameters involve smoothed probabilities which had not been computed in Shumway and Stoffer (1991). In this paper, we developed new methodology to incorporate approximate smoothing into the EM algorithm.

We noted that the cost of dealing with state space models with Markov switching is high in computational effort since the possible dependent paths must be traced back through the chain. Thus, it was necessary to make some approximations in order to obtain a computationally feasible algorithm for estimation. One primary objective in this paper was to investigate the accuracy of the filtering and smoothing approximations. To this end, we developed an MCEM

to obtain more exact solutions for the smoothed states or state probabilities. We then compared the two methodologies, AEM and MCEM, to assess the accuracy of the approximations. The comparisons included comparing (1) approximate smoothing and exact smoothing (2) parameter estimates from AEM and MCEM. From the simulation studies, we can see that the approximations of the AEM are quite accurate compared to the more exact solutions obtained from MCEM. In addition, smoothers from these two methods are almost identical. The AEM converges faster than MCEM, and MCEM is more computer intensive than the AEM. In closing we recommend employing the approximation-based algorithm, because the benefit of using it is clear—it not only leads to a significant reduction in the computational costs but also provides accurate results.

We have applied our methodology to model biological time series with changing regimes. The model has the flexibility to deal with a wide range of problems that involve possible regime shifts in pattern. The sorts of phenomena present in the influenza mortality series may be seen to occur in numerous morbidity and mortality series in medical and epidemiological studies. Our model is general enough to include most structural models useful in monitoring changes of regime.

Appendix A. Estimation of hyperparameters (EM)

For the state space model (1) with Markov switching, we consider the complete data likelihood function

$$L_{x_1, \dots, x_n, y_1, \dots, y_n, s_1, s_2, \dots, s_n}(\Theta) = f(x_1, \dots, x_n, y_1, \dots, y_n, s_1, s_2, \dots, s_n).$$

We can show that

$$\begin{aligned} -2 \ln L_{x_1, \dots, x_n, y_1, \dots, y_n, s_1, s_2, \dots, s_n}(\Theta) &\propto \ln |\Sigma_0| + (x_0 - \mu_0)' \Sigma^{-1} (x_0 - \mu_0) \\ &+ n \ln |Q| + \sum_{t=1}^n (x_t - \phi x_{t-1})' Q^{-1} (x_t - \phi x_{t-1}) - 2 \sum_{j=1}^M I_j(1) \ln \pi_j - 2 \sum_{t=2}^n \left(\sum_{i,j=1, \dots, M} I_{ij}(t) \ln \pi_{ij} \right) \\ &+ n \ln |R| + \sum_{t=1}^n \sum_{j=1}^M I_j(t) (y_t - A_j x_t)' R^{-1} (y_t - A_j x_t), \end{aligned}$$

where $\Theta = \{\mu_0, \Sigma_0, \phi, Q, R, \pi\}$, $\pi = \{\pi_j = P(s_1 = j), \pi_{ij} = P(s_t = j | s_{t-1} = i), i = 1, \dots, n, j = 1, 2, \dots, n\}$, and $I_j(t)$ and $I_{ij}(t)$ are the indicator functions of $s_t = j$ and $s_t = j, s_{t-1} = i$, respectively.

Note that $E(I_j(t) | Y_n) = \pi_j(t|n)$ and $E(I_{ij}(t) | Y_n) = \pi_{ij}(t|n)$ and that, since given Y_n, s_t and $y_t - A_j x_t$ are independent, we have

$$\begin{aligned} E(I_j(t)(y_t - A_j x_t)' R^{-1} (y_t - A_j x_t) | Y_n) \\ = E((y_t - A_j x_t)' R^{-1} (y_t - A_j x_t) | Y_n) E(I_j(t) | Y_n) \\ = \text{tr}\{R^{-1}((y_t - A_j x_{t|n})(y_t - A_j x_{t|n})' + A_j P_{t|n} A_j') \Pr(s_t = j | Y_n)\}. \end{aligned}$$

Therefore, at iteration l ,

$$\begin{aligned} E(-2 \ln L_{x_1, \dots, x_n, y_1, \dots, y_n, s_1, s_2, \dots, s_n}(\Theta) | Y_n, \Theta^{(l-1)}) \\ \propto \ln |\Sigma_0| + \text{tr}\{\Sigma_0^{-1}(P_{0|n} + (x_{0|n} - \mu_0)(x_{0|n} - \mu_0)')\} \\ + n \ln |Q| + \text{tr}\{Q^{-1}(S_{11} - S_{10}\phi' - \phi S_{10}' + \phi S_{00}\phi')\} \\ - 2 \sum_{j=1}^M \pi_j(1|n) \ln \pi_j - 2 \sum_{t=2}^n \left(\sum_{i,j=1,2, \dots, M} \pi_{ij}(t|n) \ln \pi_{ij} \right) + n \ln |R| \\ + \text{tr} \left\{ R^{-1} \left(\sum_{t=1}^n \sum_{j=1}^M (y_t - A_j x_{t|n})(y_t - A_j x_{t|n})' + A_j P_{t|n} A_j' \right) \pi_j(t|n) \right\}. \end{aligned} \quad (\text{A.1})$$

Minimizing each term in (A.1) yields the following updated estimators:

$$\hat{\pi}_j = \pi_j(1|n), \quad (\text{A.2})$$

$$\hat{\pi}_{ij} = \frac{\sum_{t=2}^n \pi_{ij}(t|n)}{\sum_{t=2}^n \Pr(S_{t-1} = i|Y_n)}, \quad (\text{A.3})$$

$$\hat{x}_0 = x_{0|n}, \quad (\text{A.4})$$

$$\hat{\phi} = S_{10}S_{00}^{-1}, \quad (\text{A.5})$$

$$\hat{Q} = n^{-1}(S_{11} - S_{10}S_{00}^{-1}S_{10}'), \quad (\text{A.6})$$

$$\hat{R} = \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^M [(y_t - A_j x_{t|n})(y_t - A_j x_{t|n})' + A_j P_{t|n} A_j'] \pi_j(t|n), \quad (\text{A.7})$$

where

$$S_{00} = \sum_{t=1}^n (x_{t-1|n} x_{t-1|n}' + P_{t-1|n}),$$

$$S_{10} = \sum_{t=1}^n (x_{t|n} x_{t-1|n}' + P_{t,t-1|n}),$$

$$S_{11} = \sum_{t=1}^n (x_{t|n} x_{t|n}' + P_{t|n}).$$

To implement the EM algorithm, we need to incorporate the smoothed probabilities $\Pr(s_t = j, s_{t+1} = k|Y_n)$ into S_{00} , S_{10} , and S_{11} via (13)–(15). Finally, we calculate $x_{0|n}$ and $P_{0|n}$, using

$$x_{0|n} = x_{0|0} + J_0(x_{1|n} - x_{1|0})$$

and

$$P_{0|n} = P_{0|0} + J_0(P_{1|n} - P_{1|0})J_0'.$$

References

- Carlin, B.P., Polson, N.G., Stoffer, D.S., 1992. A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *J. Amer. Statist. Assoc.* 87, 493–500.
- Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Engle, C., Kim, C.-J., 1999. The long-run U.S./U.K. real exchange rate. *J. Money Credit Banking* 31, 335–356.
- Kim, C.-J., 1993. Unobserved-component time series models with Markov-switching heteroscedasticity. *J. Business Econom. Statist.* 11, 341–349.
- Kim, C.-J., 1994. Dynamic linear models with Markov switching. *J. Econometrics* 60, 1–22.
- Kim, C.-J., Nelson, Ch.R., 1999. *State Space Models with Regime Switching*. MIT Press, Cambridge, MA.
- Neal, R., Hilton, G., 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. In: Jordan, M. (Ed.), *Learning in Graphical Models*.
- Pena, D., Guttman, I., 1988. A Bayesian approach to robustifying the Kalman filter. In: Spall, J.C. (Ed.), *Bayesian Analysis of Time Series and Dynamic Linear Models*. Marcel Dekker, New York, pp. 227–254.
- Shumway, R.H., 1988. *Applied Statistical Time Series Analysis*. Prentice-Hall, New York.
- Shumway, R.H., Katzoff, M.J., 1992. Nonlinear and non-Gaussian modeling of mortality series. Technical Report No. 240, Department of Statistics, University of California, Davis, CA.
- Shumway, R.H., Stoffer, D.S., 1982. An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.* 3, 253–264.
- Shumway, R.H., Stoffer, D.S., 1991. Dynamic linear models with switching. *J. Amer. Statist. Assoc.* 86, 763–769 (correction: V87 p. 913).
- Shumway, R.H., Stoffer, D.S., 2000. *Time Series Analysis and its Applications*. Springer, New York.
- Smith, A.F.M., West, M., 1983. Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics* 39, 867–878.